Dr. Cullen Jackson

Development of Human Performance Measures for Analyzing Design Issues in Submarine Tactical Control Systems

ABSTRACT

For submarine combat control, the introduction of new tools and displays has the potential to change the most effective processes for the Fire Control division. We propose a unique tool called DIVES that will facilitate a rigorous evaluation of the submarine Combat Control System (CCS) by Navy engineers, evaluators, and Fire Control Technicians. DIVES focuses on the collection of observer-based measurements of system performance based on system use by an operator, and the correlation of this data with contextual and outcome data from the Combat Control System itself. Within the laboratory setting, by finding where errors occur using the system data, system engineers will be able to correlate these errors with issues found by observing the use of the system by an operator. Subsequent analyses will drive recommendations as to the redesign of the human-machine interactions and possibly the collaboration between different positions both within and without the Fire Control division.

INTRODUCTION

The U.S. Navy is taking advantage of the revolution in information technology to support rapid evolution of its command and control systems. Along with the ability to rapidly reconfigure displays and controls, however, comes the need to rigorously evaluate the effectiveness of current and emerging human-machine interface designs. This evaluation must produce information not only on the usability of designs but also on their effectiveness in supporting mission success.

The Navy currently does not have the necessary tools to evaluate the impact of new submarine Tactical Control Systems (TCS) on human performance and decision making (Guertin and Kalisz, 2004). This need is critical to Navy operations as understanding how systems affect user performance can help reduce errors and improve accuracy and reaction times. A systematic evaluation of the effects of human-machine interface designs on mission performance requires the acquisition, integration, and use of several data sources: (1) information on an operator's interaction with the system, which can include keystroke or clickstream data; (2) information on the events occurring in the battlespace that set the context for use; and (3) information on the interactions that are occurring within a combat team or between teams that may not be captured by keystroke data, such as verbal communications and team coordination; the integration of these different data streams to produce an accurate picture of the effects of the technology on the operator, and subsequently on mission performance.

The human factors community has long understood that the introduction of new technology changes the nature of the work performed. For submarine Tactical Control Systems, the introduction of new user interfaces has the potential to change the most effective processes for the Fire Control division, to enable or even require new positions and/or responsibilities, and to potentially change the number of individuals needed within a division and the skills and knowledge they require. Data collection for evaluating the performance effectiveness for new or existing systems also should provide insight into how the systems affect the work processes both within the team and between teams in the Control Room.

In work performed for the U.S. Navy, we researched the current processes for evaluating Tactical Control Systems by the Naval Undersea Warfare Center (NUWC) and the needs of system evaluators for additional analysis capabilities. We subsequently developed a prototype software application, SPOTLITE-TCS, for collecting observer-based performance measurements of operator performance, and a prototype design for an analysis tool, called DIVES, to provide a facility for integrating multiple data sources, including observer-based data from SPOTLITE-TCS, in order to analyze which elements of the TCS need redesign and to provide recommendations for those changes.

As part of the development of the SPOTLITE-TCS tool, we developed a set of performance indicators and measures designed to allow an expert observer to watch the behaviors of a Fire Control Technician using the Tactical Control System and rate his performance. In turn, this data would allow an engineer to analyze the areas of concern with respect to the design of the operator's interactions with the TCS. In the work presented here, we outline the method used to develop the performance indicators and measures for the Fire Control Technician, introduce the SPOTLITE-TCS tool, and describe how these human performance metrics can be used within a framework to aid in the analysis and re-design of the submarine Tactical Control System to optimize operator performance and therefore mission outcomes.

DEVELOPMENT OF PERFORMANCE INDICATORS AND MEASURES

In order to evaluate the performance of a system, one must evaluate all aspects of the system including its effectiveness in supporting optimal human performance. Though one can easily obtain descriptive, outcome-based measures of human performance, e.g., success rates, time to task completion, etc., these measures are not diagnostic. As such, they do not speak to the effectiveness of the system design in supporting human performance. Diagnostic measures of human-system performance require a pluralistic approach. A performance diagnosis should use all available information, including measures, assessments, mission context, outcome data, and possibly performance of other team members in order to identify the system-based cause of the performance decrement. This analysis also might include using historical performance measurement, assessment, and outcome data captured for the same systems being used in similar situations and scenarios.

In a previous experiment (Operator Effectiveness Testing) with operators using the TCS, NUWC researchers attempted to characterize system performance by assessing operator performance. During this testing, the researchers collected multiple types of data including: system data, observer-based data, observations from the participants, and a post-event questionnaire. The observer-based data was captured using the FIT-System (Flexible Interface Technique) (Held, Bruesch, Krueger, and Pasch, 1999; Held and Manser, 2005). With this tool, a template is created that approximately replicates the layout of the controls on the TCS display, and this template is fit over the screen of a personal data assistant (PDA). During the experiment, the observer selected the part of the system being used or attended to by clicking on the template. Most of the analyses were based only on system data and, without other data sources to corroborate the system outcome data or a priori metrics for indicating operator performance, it was difficult to diagnose the system issues that contributed to poor operator performance. In addition, while this was not the focus of the experiment, there existed no facility for collecting data on interactions either within Fire Control or between teams in the Control Room.

Development of Fire Control Performance Indicators

The development of performance indicators and subsequent measures provide a theory-based, a priori framework for determining operator and team performance. Performance indicators (PIs) can provide information on what operator and team behaviors constitute poor or good performance during Fire Control operations (e.g., "Recognizes contact on a constant bearing"). Aptima has developed processes for creating performance indicators and measures and technologies designed to utilize them over the past several years: the methodology for creating performance indicators and measures is a process we call COMPASS (Competency-based Measures for Performance ASsessment Systems) and our observer-based measurement technology is called SPOTLITE (Scenario-based Performance Observation Tool for Learning In Team Environments). Figure 1 illustrates the relationship between the COMPASS process for creating performance indicators and measures and the SPOTLITE tool for actually implementing and taking those measures.

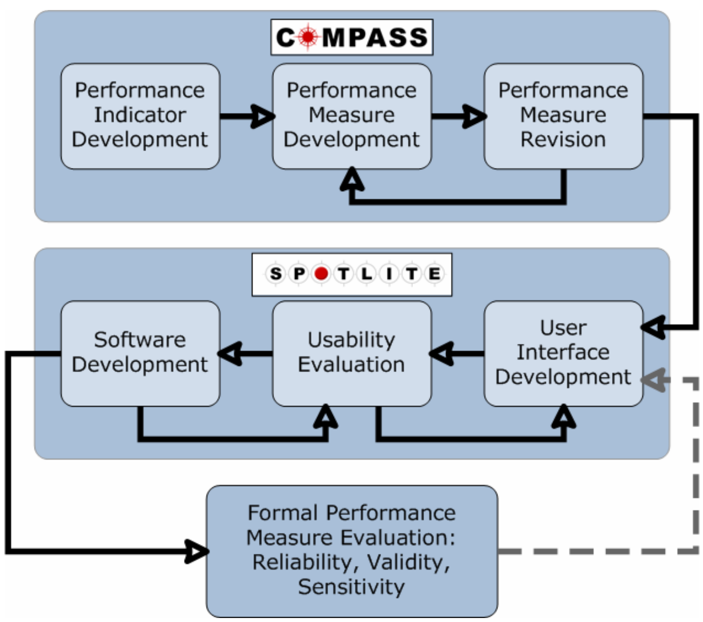


Figure 1. Relationship between the COMPASS process and the SPOTLITE tool.

A performance indicator is an observable behavior that allows an expert to rate an operator's performance in real-time; in this case Fire Control Technician performance using the TCS. Performance indicators are elicited from subject matter experts (SMEs) in order to ensure that they are accurate representations of necessary operator performance in appropriate mission contexts. Performance indicators are therefore an integral part of the system evaluation process as they ensure that measured operator performance while using a system will translate in an operational situation. The process of defining and refining performance indicators is done as part of the COMPASS process, which occurs during three separate workshops with small groups of subject matter experts in the domain of interest. Generally, performance indicators and performance measures, along with the appropriate workflow for the domain and tasks, are defined during the first and second COMPASS workshops. For this effort we were not able to constitute a single group of SMEs to participate in consolidated workshops. Instead, we obtained our performance indicators for Fire Control through several separate interview sessions with four senior Fire Control Technicians; three of our subject matter experts were rated FTC(SS) and the fourth was rated FTCS(SS) and a senior Sonar Technician (STSCS). We interviewed the senior Sonar Technician in order to ensure that we understood the tasks involved in Fire Control from multiple perspectives.

During these interviews, we asked our SMEs to indicate operator performance associated with using the TCS that they had frequently observed as expert users and supervisors of TCS users. During this process, we focused exclusively on behaviors that the SMEs had observed and that they felt we could observe during system evaluations rather than those behaviors that could be inferred. Through this process, we ensured that we captured those behaviors that would best serve as a basis on which to develop performance measures that are less sensitive to individual rater differences and can be reliably rated by multiple observers (MacMillan, Entin, and Morley, in press). A subset of performance indicators developed as a result of this process is shown in Table 1; these performance indicators are quite broad and they would be refined into more detailed indicators in subsequent working groups.

Task	Performance Indicator (PI)	UI or Team Design Issue
L	Deals appropriately with unreliable contact information	Reliability or uncertainty of contact information is not discernable through TCS displays or tools
T	Performs contact evaluation in a timely manner (5 mins/contact)	TCS menu structure is too dense and does not facilitate rapid interaction with the system
T	Uses algorithm solutions to refine own system solution	Algorithm displays do not afford rapid or easy interpretation of their solutions
T	Refines system solutions based on contact priority (e.g., closest point of approach)	TCS does not provide enough information to assess the contacts of highest priority
T	Evaluates solution based on multiple sources of information	TCS does not allow for the display of multiple tools that may be used for solution validation

Table 1. Set of performance indicators for Fire Control and their mapping to TCS design issues. In the Task column, L stands for “Localize” and T stands for “Track.”

As previously described, in order for the Fire Control performance indicators to inform the researchers about the relationship between operator performance and system design issues, the performance indicators must be mapped to possible system design problems. For this mapping, we consulted human factors subject matter experts to develop a mapping of the performance indicators to system design issues (see the right-most column in Table 1).

Development of Performance Measures

After development of the Fire Control performance indicators, the next step in the COMPASS process is to develop the set of associated performance measures. This process involves taking the performance indicators, which are based on observable behaviors, and creating a measure with associated behavioral anchors that define good and poor performance. These measures can describe observable behaviors for both individuals and teams. This capability will allow system evaluators to capture data on team interactions and team performance with possible mappings to recommendations that will provide for increased team interaction and coordination by improving the design of associated system elements. Specific questions that helped identify these observable behaviors and associated anchors for Fire Control and TCS use included:

- What does an operator do or say to indicate good/poor performance for this performance indicator?
- What would cause an operator to do well or poorly at this performance indicator?
- In what situations will a person perform well or poorly for this performance indicator?
- What specific TCS tools do they use to help them accomplish this performance indicator?

For this effort, we concentrated on developing performance measures for those performance indicators that were readily translated into observer-based performance measures, i.e., the performance indicators identified by SMEs with explicit behaviors representative of good or poor performance. Figure 2 shows three performance measures developed for observing Fire Control Technician performance while using the TCS.

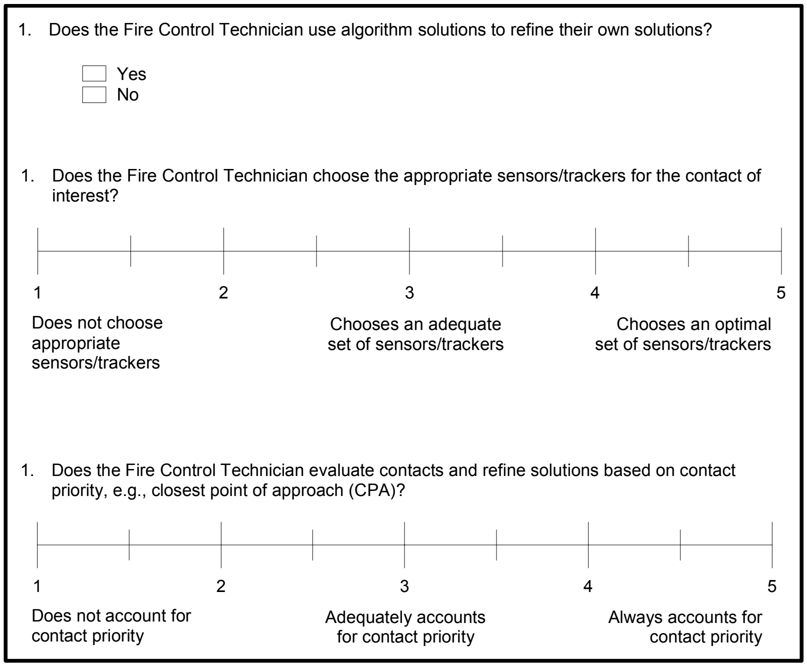
DEVELOPMENT OF THE SPOTLITE-TCS TOOL

In developing the prototype version of the SPOTLITE-TCS tool, we leveraged previous work on SPOTLITE for the Air Force Research Laboratory, specifically our work developing SPOTLITE DTC, a module for SPOTLITE focused on performance measurement for the Dynamic Targeting Cell of the Air and Space Operations Center (Morley et al., 2005). We chose this implementation of SPOTLITE as a template because it is our only module focused on a command and control operational element and therefore our closest analogy to submarine Fire Control. Overall, our goal was to implement SPOTLITE-TCS such that it allowed one observer to rate one Fire Control Technician at a time in real-time, rate multiple contacts during one session, and provided "hooks" for future functionality and integration.

Specifically, the SPOTLITE-TCS design supports the following:

- Quick and easy *real-time* assessment
- Increased **usability** for expert observers
  - Clear measure wording
  - Clear visibility of rating progress using progress bars
- Intuitive interaction with SPOTLITE-TCS
  - Supports Fire Control task workflow and processes
- Clickable Likert and bi-variate scales for indicating measurements

One of the primary carryovers from SPOTLITE DTC is the use of a familiar workflow to anchor the performance measures and support scenario flow and expert observer expectations. In the case of SPOTLITE DTC, this workflow was the F2T2EA kill chain while for SPOTLITE-TCS we used the five task categories developed as part of the COMPASS process: Search, Localize, Track, Engage, Analysis.



1. Does the Fire Control Technician use algorithm solutions to refine their own solutions?

☐ Yes
☐ No

1. Does the Fire Control Technician choose the appropriate sensors/trackers for the contact of interest?

1 2 3 4 5

Does not choose appropriate sensors/trackers
Chooses an adequate set of sensors/trackers
Chooses an optimal set of sensors/trackers

1. Does the Fire Control Technician evaluate contacts and refine solutions based on contact priority, e.g., closest point of approach (CPA)?

1 2 3 4 5

Does not account for contact priority
Adequately accounts for contact priority
Always accounts for contact priority

Figure 2. Three measures based on performance indicators eliciting during the COMPASS process.

The main screen of the SPOTLITE-TCS tool allows the observer to add and select contacts using a pull-down menu, view their progress in filling in measures using progress bars (for the contact and overall scenario), the ability to jump between tasks using the Fire Control workflow on the left-hand side, to add comments by clicking on the pencil icons and entering text, and of course to perform measures.

ANALYTIC FRAMEWORK

Aptima and the Center for Intelligent Information Retrieval at the University of Massachusetts – Amherst have partnered to develop UPDATE, a system designed to classify new and existing usability reports and provide an analysis interface for diagnosing usability faults and identifying trends in the data. Part of this project created a common conceptual scheme – taxonomy of usability problems – to provide analysts with a shared perspective on the design-centered roots of behaviors observed and a common language that supports analysis across individuals, studies, and organizations. The goal of UPDATE is to help usability analysts observe and document user behaviors, diagnose design faults from observational data, and develop design solutions that are optimized both locally and globally. For this effort, we planned modify UPDATE to accept system-based outcome data in place of the usability reports generally input to the UPDATE tool; to date, we have not made these modifications.

The revised UPDATE system will classify this data in accordance with the observer-based performance measures and descriptions by the observers on TCS use. The system will provide an interface for diagnosing system faults and for identifying trends in the TCS performance data. The UPDATE taxonomy will still provide the ability for analysts to view a shared perspective on the design-centered roots of behaviors observed to support analysis across individuals, teams, and studies/experiments; this capability will enable researchers to develop system re-design solutions for the TCS that are optimized both locally and globally. Figure 3 illustrates the process flow within the UPDATE tool, which will serve as the analytical backend of the DIVES system. The observer-based performance data collected via the SPOTLITE-TCS tool would be integrated into the UPDATE process through the "Describe the symptoms of the use issue – observation" module, while issues derived from data directly from the CCS would be input through the "Detect performance difficulties" module.

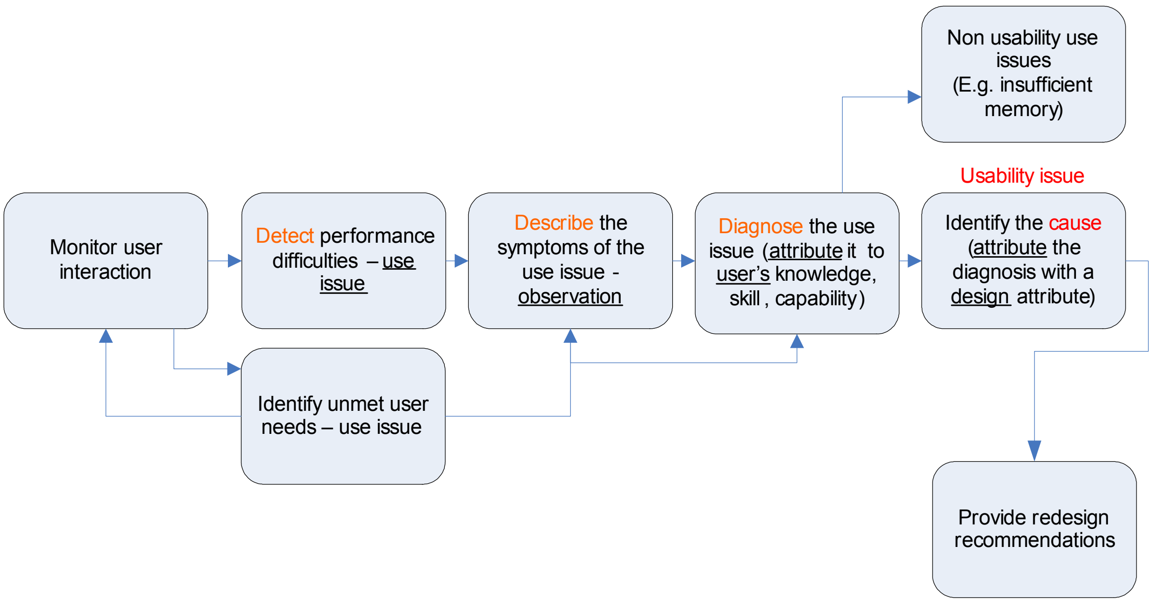


Figure 3. Schematic of the process flow within the UPDATE system.

CONCLUSION

Situational awareness is often cited in the military as a prerequisite to efficient mission performance. Warfighters require detailed information about the state of essential mission elements – own forces, targets, weather conditions – to make accurate assessments and effective decisions. Display systems for warfighter situation awareness are being upgraded and redesigned at a very high rate as legacy systems are being replaced by commercial-off-the-shelf systems. Along with the ability to rapidly reconfigure displays and controls, however, comes the need to rigorously evaluate the effectiveness of new human-machine interface designs. This evaluation must produce information not only on the usability of designs but also on their efficacy in supporting mission success.

We have described the initial work to develop a framework that will allow for the integration of system-based and observer-based metrics to inform the re-design of submarine Tactical Control Systems (TCS). This work involved developing performance indicators and measures for the Fire Control Technician, developing a prototype TabletPC-based performance measurement tool (SPOTLITE-TCS), and designing a framework within which these human performance metrics can be used to aid in the analysis and re-design of the submarine Tactical Control System to optimize operator performance and therefore mission outcomes. This unique tool, DIVES, will facilitate a rigorous evaluation of the submarine TCS by Navy researchers, evaluators, and Fire Control Technicians.

REFERENCES

Guertin, N.H. and M.S. Kalisz, “Submarine Combat Control System Development With a Focus on Human Systems Integration,” *Proceedings of the Undersea Human/Systems Integration Symposium*. Newport, RI: Naval Undersea Warfare Center, Division Newport (NUWCDIVNPT), May 2004.

Held, J., M. Bruesch, H. Krueger, and T. Pasch, “The FIT-System: A New Hand-held Computer Tool for Ergonomic Assessment,” *Medical and Biological Engineering and Computing, Suppl. 2 Vol. 37,* pp. 862-3., 1999.

Held, J. and T. Manser, “A PDA-based System for Online Recording and Analysis of Concurrent Events in Complex Behavioral Processes,” *Behavioral Research Methods, 37(1),* pp. 155-64, 2005.

MacMillan, J., E.B. Entin, R. Morley, “Measuring Team Performance in Complex and Dynamic Military Environments: The SPOTLITE Method,” *Military Psychology*, in press.

Morley, R., E. Entin, M. Garrity, L. Beer, C. Tossell, “Developing a Performance Measurement Instrument for Team Performance in the Air and Space Operations Center,” *Proceedings of the Training, Education, and Simulation International Conference*, 2005.

ACKNOWLEDGMENTS

I’d like to thank Dr. Amy Alexander, Dr. Rebecca Grier, Susannah Hoch, and Emily Wiese for their work on this project, and Daniel Lawrence for providing key insights into the inner workings within the control room of a U.S. Navy submarine. I’d also like to thank Dr. Joseph Gabriel, Megan Gibson, Dr. Susan Kirschenbaum and several of the Fire Control Chiefs at NUWC-Newport for their support during this effort. The work described in this manuscript was performed under U.S. Navy contract N65538-05-M-0174, and the views contained herein have not been endorsed by the U.S. Navy, Department of Defense, or any other part of the U.S. Government.

Dr. Cullen Jackson *is the Performance Assessment Area Lead at Aptima, Inc. Dr. Jackson applies his knowledge of human perceptual and cognitive capabilities to designing systems and techniques for measuring the impact of technology on human performance effectiveness. Dr. Jackson also applies his expertise to simulator- and computer-based training systems, information and uncertainty visualization systems, user interfaces and tactical displays, and communications assessment techniques. Dr. Jackson holds a Ph.D. in Experimental Psychology from Brown University and a B.S. in Computer Science from Trinity University. He is a member of the Cognitive Science Society, the Association for Computing Machinery, and the IEEE Computer Society.*